DOCUMENT RESUME

ED 229 414 TM 830 313

AUTHOR Wolf, Fredric M.; And Others
TITLE Predictive and Incremental Validity of the New MCAT Science Problems Subtest.
PUB DATE Apr 83
NOTE 19p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Montreal, Quebec, April 12-14, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Entrance Examinations; *Competitive Selection; Higher Education; *Medical Students; Predictive Measurement; *Predictive Validity; Regression (Statistics); Science Tests; *Testing Problems; Test Items
IDENTIFIERS Cross Validation; *Medical College Admission Test; *National Board of Medical Examiners; Test Revision

ABSTRACT

The predictive and incremental validity of the New Medical College Admission Test (New MCAT) Science Problems Subtest was examined with a sample of over 165 medical students. Criterion measures were National Board of Medical Examiners (NBME) Part I (basic science) and Part II (clinical science) performance. The Science Problems subscore is derived from a subset of the same items found on the Biology, Chemistry, and Physics subtests, creating nonindependence. Results of incremental validity analyses and of all possible subsets regression analyses using Mallow's Cp criterion raise questions concerning the practical utility of the Science Problems subtest in prediction equations to make admission decisions. Cross-validation analyses supported the inclusion of the Biology subtest in prediction models of both NBME Parts I and II, and of the Chemistry subtest for NBME Part I. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED229414

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Predictive and Incremental Validity of the New MCAT Science Problems Subtest

Fredric M. Wolf, Judith G. Calhoun,
Bruce R. Maxim, Wayne K. Davis

University of Michigan
Medical School

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

F. M. Wolf

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Running Head: Science Problems

Presented at the meeting of the National Council on Measurement in Education, Montreal, April 1983. Correspondence may be addressed to the authors at the Department of Postgraduate Medicine and Health Professions Education, University of Michigan. G1208 Towsley Center, Ann Arbor, Michigan 48109.

TM 830 33

Predictive and Incremental Validity of the New MCAT Science Problems Subtest

Abstract

The predictive and incremental validity of the New Medical College Admission Test (New MCAT) Science Problems Subtest was examined with a sample of over 165 medical students. Criterion measures were National Board of Medical Examiners (NBME) Part I (basic science) and Part II (clinical science) performance. The Science Problems subscore is derived from a subset of the same items found on the Biology, Chemistry, and Physics subtests, creating nonindependence. Results of incremental validity analyses and of all possible subsets regression analyses using Mallows C_p criterion raise questions concerning the practical utility of the Science Problems subtest in prediction equations to make admission decisions. Cross-validation analyses supported the inclusion of the Biology subtest in prediction models for both NBME Parts I and II, and of the Chemistry subtest for NBME Part I.

Predictive and Incremental Validity of the New MCAT Science Problems Subtest

The Association of American Medical Colleges revised standardized test designed to evaluate the academic preparation of applicants to medical school was first used in 1978. This version, the New Medical College Admissions Test (New MCAT) differs in several respects from the Old MCAT. "Specifically, the Skills Analyses and Science Problems subtests of the New MCAT assess such abilities as information gathering and analysis, discerning and formulating relationships and other problem solving skill dimensions in their respective areas. These cognitive areas were not directly measured by the Old MCAT" (New MCAT Interpretive Manual, 1977). Dawson-Saunders and Doolen (1981) and Jones and Thomae-Forgues (1981) discussed the New MCAT's potential value as a predictor of clinical performance. Due to the increased emphasis on interpretation and problem solving in the new format, they suggest that the new MCAT may result in measures which are more closely associated with the information gathering, evaluation, and utilization skills required during the clinical experience. A number of studies have compared the ability of the Old and New MCAT to predict student achievement in medical school. Erdmann (1980) characterized the results of the "first round" of New MCAT studies as encouraging.

Because the scores on the Science Problems subtest are derived from a subset of the items that comprise three other New MCAT subtests, Biology, Chemistry, and Physics, this subtest is by definition linearly dependent upon these other subtests. Thus, while "scores on the six New MCAT areas of assessment are designed to be relatively independent and are purposefully reported seperately. . . . items from the Science Problems subtest contribute twice to New MCAT scores" (New MCAT Interpretive Manual, 1977). This issue has been addressed in several New MCAT validity studies (Hull,

Calhoun & Maxim, 1981; Jones & Thomae-Forgues, 1981) by excluding the Science Problems subtest from multivariate analyses, while it has been included in other studies (Friedman & Bakewell, 1980; Friedman & Porter, 1981; McGuire, 1980; Molidor & Elstein, 1979). Psychometrically the problem is that the Science Problems subtest partakes of the same error component of the other subtests, violating the assumption of uncorrelated error variance, raising serious interpretative questions in multivariate analyses such as factor analysis (Gorsuch, 1974). When independent variables such as these are highly correlated in multiple regression analyses, "not only do the estimated regression coefficients tend to be quite imprecise, but the true regression coefficients tend to lose their meaning" (Neter & Wasserman, 1974). On the other hand, multicollinear variables have been included in the same analyses when strong rationale for their inclusion has been given. In a recent re-examination of the relevance of MCAT science content, neither the Science Problems subtest nor this issue of non-independence was discussed (Wilson, 1982). It is likely that the Science Problems subtest has been included in prediction equations used to make admission decisions at many medical schools. The purpose of the present study was to examine the usefulness of the New MCAT Science Problems subtest in predicting medical student basic and clinical science performance.

Methodology

Instrumentation and Sampling

Scores for the entering class of 1978 at a large midwestern University medical school were obtained for student performance on the six New MCAT subtests (Biology, Chemistry, Physics, Science Problems, Skills Analysis: Quantitative, Skills Analysis: Reading), and the examinations of the National Board of Medical Examiners (NBME), Parts I and II. NBME scores (NBME, 1982) represent the criterion medical school performance measures examined in the study. Part I assesses basic science achievement, while NBME Part II assesses clinical science achievement.

Subjects were medical students in the 1982 graduating class at The University of Michigan Medical School. Because of missing data, total sample sizes were 186 subjects for the NBME Part I analyses and 167 subjects for the NBME Part II analyses. Subjects were randomly divided into two sub-samples, a screening sample and a calibration sample, in order to cross-validate the results obtained in the multiple correlation/regression analyses (Kerlinger & Pedhazur, 1973; Lord & Novick, 1968) described in the following section. All data were analyzed for each sub-sample independently and again for the total combined sample.

Capitalization on chance in the development of a regression/prediction model based on sample correlations is a well known problem (Lord & Novick, 1968). Because these sample correlations are based not only on true correlation among the variables, but also contain sampling error, the multiple correlation typically "shrinks" when these variables are used on a new sample. Both Lord and Novick (1968) and Kerlinger and Pedhazur (1973) recommend cross-validation procedures to address this problem. Cross-validation necessitates obtaining two samples. The first sample is referred to as the screening sample, and is used to develop the regression equation and multiple R^2. The predictor variables of the second sample, referred to as the calibration sample, are then applied to the regression equation obtained from the screening sample to obtain predicted scores for the criterion variable. The observed criterion scores (y) for the calibration sample are then correlated with the predicted criterion scores (y'). This Pearson $r_{yy'}$ is analogous to a multiple correlation between the observed and predicted scores. In the present study, this procedure was applied twice in order to allow each sub-sample to constitute the screening (and calibration) sample. This "double cross-validation procedure is strongly recommended as the most rigorous approach to the validation of results from regression analysis in a predictive framework" (Kerlinger and Pedhazur, 1973, p.284). Results of the two regression equations, multiple R^2s and $r_{yy'}$s obtained from alternate samples were then compared. Analyses of the data were performed retrospectively and were not used in making admission decisions.

Correlational and Incremental Validity Analyses

Pearson zero order correlations were computed to test the research hypotheses of a significant positive relationship between each of the MCAT subscores and the two criterion performance measures. Incremental validity (Lord & Novick, 1968; Sechrest, 1963) was examined by using a step-wise, hierarchical multiple regression analysis design involving a two step procedure. In the first phase, all MCAT subtest scores except Science Problems were included in the analysis. The Science Problems subscores were then included in the second phase of the analysis by stepping them into the equation after the non-Science Problems Subtest had been stepped in. Two seperate analyses were performed, one for each of the criterion measures. These analyses permitted an examination of the usefulness of the Science Problems subtest in explaining additional variance in the criterion measures beyond that already explained by the other MCAT subtests.

Three separate indexes of MCAT incremental validity were calculated. The first index indicates the absolute amount of variance (as measured by multiple R^2) explained for each of the two criterion measures by the Science Problems subtest scores when they are stepped into the multiple regression analysis after all the non-Science Problems MCAT subtests have been included (Sechrest, 1963). This index was determined using formula 1.

Index 1= (R^2 for all variables) - (R^2 for non-Science Problems MCAT variables) (1)

= R^2 added by Science Problems MCAT Subtest

The second index (Friedman & Porter, 1981) provides a measure of the proportional increase in performance variance explained by stepping in MCAT Science Problems

subtest scores last in the regression analysis and was calculated using formula 2 below.

$$\text{Index 2} = \frac{R^2 \text{ added by Science Problems MCAT}}{R^2 \text{ for non-Science Problems MCAT variables}} \quad (2)$$

The third index provides a measure of the proportional increase in performance variance that is unaccounted for by the non-Science Problems MCAT subtests and that is explained by adding the Science Problems MCAT scores to the regression analysis (Friedman & Porter, 1981). This index was calculated using formula 3 below.

$$\text{Index 3} = \frac{R^2 \text{ added by Science Problems MCAT}}{1 - (R^2 \text{ for non-Science Problems MCAT variables})} \quad (3)$$

Both indexes 2 and 3 are calculated in order to minimize artifactual differences in the incremental validity results (Freidman & Porter, 1981) of the Science Problems for the two samples due to differing multiple R^2s or differing amounts of unexplained variance available (i.e., not explained by the non-Science Problems subtests).

All Possible Subsets Regression Analyses

All possible subsets regression analyses (Frane, 1981) including all six New MCAT subtests are reported for each of the criterion measures. "The only way to be sure of obtaining the best n of N predictors would be to determine the multiple correlation for every such set" by using an exhaustive procedure (Lord & Novick, 1968, p. 288). Until recently the economic cost of performing such analyses was prohibitive. However, "one major advance of the past decade in multiple regression has been the replacement of

stepwise procedures with all possible subset searches for model selection, served by the C_p plot "(Wainer & Thissen, 1981, p. 213). Use of the Furnival-Wilson (1974) algorithm enables the identification of "subsets while computing only a small fraction of all possible regressions. Computer costs are comparable for stepwise regression for up to about 25 independent variables" (Frane, 1981, p. 264).

Mallow's C_p was the criterion used to identify the best subsets. The "best" subset is selected on the basis of an analysis of residuals that minimizes C_p based on the following formula (Daniel & Wood, 1971; Frane, 1981):

$$C_p = \frac{RSS}{s^2} - (N-2p') \qquad (4)$$

where

RSS= residual sum of squares for the subset of independent variables being tested

s^2= residual mean square based on the regression using all independent variables

p'= the number of variables in the subset, including the intercept, if any.

n= number of cases (sample size)

In addition, multiple R^2s and adjusted R^2s based on formula 5 were calculated.

$$\text{Adjusted } R^2 = \frac{R^2 - p(1-R^2)}{N - p'} \qquad (5)$$

where p= the number of independent variables when the intercept is set to zero.

These analyses enabled an examination of which subtests, the Science Problems subtest and/or other subtests, were included in the "best" regression model for each criterion.

Results and Discussion

Intercorrelations among all six New MCAT subtests and NBME Part I scores are summarized in Table 1 for both subsamples. Similar correlations are presented in Table 2 for NBME Part II analyses. Table 3 contains intercorrelations for all subjects (i.e, both subsamples combined). The colinearity was greatest for the relationships of the Science Problems Subtest with the Biology, Physics and Chemistry subtests (r's ranged between .56 and .72). Not surprisingly, the content non-independence resulting from the use of selected items from the Biology, Physics, and Chemistry subtests in the construction of the Science Problems subtest is confirmed by the magnitude of these correlations. Because the largest amount of shared variance is 52%, it could be argued that there is sufficient non-overlap to justify the inclusion of the Science Problems subtest on theoretical grounds. Correlation coefficients involving Science Problems with the NBME measures were exceeded in magnitude only by the Biology subtest correlation coefficients, except in subsample 2 where the Science Problems - NBME Part I correlation actually exceeded the Biology correlation (.38 versus .35). These Pearson correlations between Science Problems and NBME measures ranged between .38 and .55. Correlation coefficients in general were higher in sample 1 than in sample 2. This could possibly reflect greater variability among NBME Part I and II scores among subjects in sample 1 than in sample 2.

Table 1

Pearson Correlations Among New MCAT Subtests and NBME Part I Scores

		Sample 1 (n=92)						
		BI	PH	CH	SP	RE	QA	NBMEI
Sample 2 (n=94)	MCAT-BI	---	.49	.48	.66	.26	.42	.62
	MCAT-PH	.31	---	.62	.65	.23	.36	.47
	MCAT-CH	.30	.52	---	.72	.31	.40	.49
	MCAT-SP	.57	.61	.59	---	.25	.44	.55
	MCAT-RE	.12	.27	.28	.33	---	.26	.30
	MCAT-QA	.25	.37	.38	.48	.25	---	.33
	NBME I	.35	.29	.35	.38	.17	.24	---

Note: MCAT= New Medical College Admission Test; BI= Biology; PH= Physics; CH= Chemistry; SP= Science Problems; RE= Reading; QA= Quantitative.

All correlations greater than .205 or .267 are statistically significant at alpha=.05 or alpha=.01, respectively (df=90).

Table 2

Pearson Correlations Among New MCAT Subtests and NBME Part II Scores

		Sample 1 (n=81)						
		BI	PH	CH	SP	RE	QA	NBMEII
Sample 2 (n=86)	MCAT-BI	---	.50	.52	.68	.28	.44	.56
	MCAT-PH	.29	---	.61	.64	.17	.37	.37
	MCAT-CH	.28	.50	---	.73	.30	.43	.36
	MCAT-SP	.58	.59	.56	---	.22	.46	.43
	MCAT-RE	.16	.27	.28	.34	---	.24	.40
	MCAT-QA	.27	.40	.38	.49	.26	---	.43
	NBME II	.37	.21	.34	.39	.21	.23	---

Note: MCAT= New Medical College Admission Test; BI= Biology; PH= Physics; CH= Chemistry; SP= Science Problems; RE= Reading; QA= Quantitative.

All correlations greater than .217 or .283 are statistically significant at alpha=.05 or alpha=.01, respectively (df=80).

Table 3

Pearson Correlations Among New MCAT Subtests and NBME Part I and II Scores for All Subjects

		NBME I (n=186)						
		BI	PH	CH	SP	RE	QA	NBME I
NBME II (n=167)	MCAT-BI	---	.41	.40	.63	.20	.34	.51
	MCAT-PH	.39	---	.57	.62	.25	.37	.38
	MCAT-CH	.41	.56	---	.66	.30	.39	.43
	MCAT-SP	.64	.61	.65	---	.28	.46	.48
	MCAT-RE	.22	.22	.29	.27	---	.26	.24
	MCAT-QA	.36	.38	.41	.47	.25	---	.29
	NBME II	.48	.29	.36	.42	.32	.35	---

Note: MCAT= New Medical College Admission Test; BI= Biology; PH= Physics; CH= Chemistry; SP= Science Problems; RE= Reading; QA= Quantitative.

All correlations are statistically significant ($p < .01$).

Incremental Validity Results

Multiple R^2s indicated that all six New MCAT subtests accounted for 45%, 20%, and 33% of the variance in NBME Part I scores for subsample 1, subsample 2, and the combined sample, respectively. In all but one of the incremental validity analyses reported in Table 4, Science Problems did not explain any additional or incremental variance in NBME measures beyond that explained by the other five New MCAT subtests. The one exception occured in Sample 2 for NBME Part II, where the multiple R^2 improved from .21 to .22 with the addition of Science Problems. Using formula 1, Science Problems explained only 1% additional variance in this instance, for a 5% proportional increase in performance variance explained (formula 2) and 1% of the variance unaccounted for

by the other New MCAT subtests (formula 3). In general, these incremental validity analyses raise doubts concerning the practical utility of the Science Problems subtest in explaining variability among NBME performance not accounted for by the other subtests.

Table 4

Incremental Validity for New MCAT Science Problems Subtest

Criterion Measure	Statistic	Sample 1	Sample 2	All Subjects
NBME I	Sample Size (n)	92	94	186
	R^2 non-Sci. Prob. MCAT	.45	.20	.33
	R^2 added by Sci. Prob. MCAT (1)	.00	.00	.00
	Total R^2	.45	.20	.33
	Incremental Validity (2)	.00	.00	.00
	Incremental Validity (3)	.00	.00	.00
NBME II	Sample size (n)	81	86	167
	R^2 non-Sci. Prob. MCAT	.41	.21	.31
	R^2 added by Sci. Prob. MCAT (1)	.00	.01	.00
	Total R^2	.41	.22	.31
	Incremental Validity (2)	.00	.05	.00
	Incremental Validity (3)	.00	.01	.00

All Possible Subsets Regression Results

These analyses were performed to examine whether the Science Problems subtest was a component of the best regression models for predicting NBME performance. Based on the selection criterion of minimizing the C_p statistic for residuals, the following standardized regression models were obtained for sample 1 (equation 6) and sample 2 (equation 7):

$$\text{NBMEI.1} = .49\ \text{Biology} + .26\ \text{Chemistry} + 1.18 \quad (6)$$

$$\text{NBMEI.2} = .26\ \text{Biology} + .28\ \text{Chemistry} + 3.13 \quad (7)$$

Even though there are differences in the beta weights between the two models, there is striking similarity between them. These results support the cross-validity and plausibility of a prediction model for NBME Part I scores that include only the Biology and Chemistry subtests. The two subsamples were then combined to provide a more stable regression equation (Kerlinger & Pedhazur, 1973; Mosier, 1951) and is presented below.

NBMEI = .39 Biology + .25 Chemistry + .0⁹ Reading + 1.44 (8)

This model was selected based on having the lowest C_p value (3.31). However, the model comprised of just Chemistry and Biology resulted in a C_p value of 3.36. Combined with Frane's (1981) recommendation that only independent variables whose coefficients are significantly different from zero be retained, it is unlikely that adding the Reading subtest would result in predictions substantially different from excluding it from the model (the beta coefficient of .09 was not statistically significant, $p < .16$).

The regression models obtained for NBME Part II performance for sample 1 (equation 9) and sample 2 (equation 10) contained both similarities and differences.

NBME II.1 = .41 Biology + .24 Reading + .20 Quantitative -35 (9)

NBME II.2 = .30 Biology + .26 Chemistry + 2.40 (10)

It seems clear that Biology is a good predictor and should be included in the model. Results for the Reading, Quantitative, and Chemistry subtests are ambiguous, as their contributions were not cross-validated. Combining both subsamples resulted in the following regression model:

NBME II = .35 Biology+.18 Reading+.13 Quantitative+.11 Chemistry+.45 (11)

Not surprisingly, the best model for all subjects included all four subtests included in equations 9 and 10. Neither the beta weights for Quantitative ($p < .08$) nor Chemistry ($p < .16$) were statistically significant. Thus while the best model for predicting NBME Part II scores is not clear based on these analyses, it is clear that Science Problems is not one of the plausible predictors under consideration.

Table 5 summarizes the C_p, multiple R^2, adjusted R^2, and $r_{yy'}$ values for the best subset regression models reported above. The $r_{yy'}$ coefficient of .66 was obtained by correlating sample 1 (calibration sample) subjects observed scores with their predicted scores based on the model derived with sample 2 (screening sample). In general, squaring the $r_{yy'}$ coefficients from each sample and comparing them with the multiple R^2 or adjusted R^2s coefficients from the same sample indicates striking similarity and consistency, particularly for NBME Part I. The difference between multiple R^2 for the two samples, as well as the difference between $r_{yy'}$ coefficients, provides an estimate of the amount of shrinkage of the multiple correlation. In general, shrinkage decreases as sample sizes increase (Kerlinger & Pedhazur, 1973). Even though the ratio of subjects to the number of independent variables ranged between 13.5:1 and 15.7:1 for the two subsamples, these samples may still be considered relatively small for the types of analyses performed. As data become available for the graduating class of 1983, it would be useful to replicate these analyses with the entire classes of 1982 and 1983 representing the two samples in contrast to dividing the class of 1982 into two subsamples as reported here.

Table 5

Mallow's C_p, Multiple R^2, Adjusted Multiple R^2, and Cross-Validated Composite Correlations ($r_{yy'}$) for Best Subset Regression Analyses

Criterion Measure	n	C_p	R^2	Adj R^2	$r_{yy'}$
NBME I					
Sample 1	92	1.90	.43	.42	.66
Sample 2	94	0.64	.19	.17	.43
All Subjects	186	3.31	.33	.31	---
NBME II					
Sample 1	81	1.79	.41	.39	.58
Sample 2	86	0.93	.20	.18	.39
All Subjects	167	3.06	.31	.29	---

Note: $r_{yy'}$ is the Pearson r "between the observed criterion scores (y) in the calibration sample and the predicted criterion scores (y'). This $r_{yy'}$ is analogous to a multiple correlation in which the equation used is the one obtained in the screening sample" (Kerlinger & Pedhazur, 1973, p. 284).

All Multiple R, Adjusted Multiple R, and $r_{yy'}$ correlations are statistically significant ($p < .001$).

Conclusions

Results of cross-validation analyses support the inclusion of Biology and Chemistry subtests in prediction models for NBME Part I performance, and of Biology for Part II performance. The contributions and utility of the Reading, Quantitative, and Chemistry subtests for predicting Part II performance are ambiguous based on the results of this study.

Both the results of incremental validity and the all possible subset regression analyses obtained in this study raise doubts concerning the usefulness of the New MCAT Science Problems subtest in predicting student performance on two widely used standardized measures of medical school basic and clinical science achievement. Combined with the psychometric issues raised in using nonindependent variables in multivariate analyses, these results suggest great care should be exercized in using the Science Problems subtest in making admission decisions. Certainly one study does not definitively resolve this issue. Replication of these findings with samples obtained from other medical schools using similar and different criterion medical school performance measures is recommended before more definitive statements are made, although the caveat from this study is clear.

References

Bulletin of Information and Description of National Board Examinations. Philadelphia: National Board of Medical Examiners, 1982.

Cohen, J. Statistical Power Analysis for the Behavioral Sciences (Rev. ed.). N.Y.: Academic Press, 1977.

Cohen, J. and Cohen, P. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Hillsdale, N.J.: Erlbaum, 1975.

Daniel, C. and Wood, F.S. Fitting Equations to Data. N.Y.: Wiley, 1971.

Dawson-Saunders, B., Doolen, D.R. An Alternative Method to Predict Performance: Canonical Redundacy Analysis. Journal of Medical Education, 56:295-299, 1981.

Erdmann, J.B. Validating the MCAT. Journal of Medical Education, 55:463-464, 1980. (editorial)

Frane, J. All Possible Subsets Regression. In W.J. Dixon and M.B. Brown (eds.). BMDP: Biomedical Computer Programs. Los Angeles: Univeristy of California Press, 264-277, 1981.

Friedman, C.P., Bakewell, Jr., W.E. Incremental Validity of the New MCAT. Journal of Medical Education, 55:399-404, 1980.

Friedman, C.P. and Porter, C.Q. Incremental Validity: The Old and New MCATs Compared. Proceedings of the 20th Annual Conference on Research in Medical Education, Washington, D.C., 251-256, 1981.

Furnival, G.M. and Wilson, R.W. Regression by Leaps and Bounds. Technometrics, 16:499-511, 1974.

Gorsuch, R.L. Factor Analysis. Philadelphia: W.B. Saunders, 1974.

Hull, A.L., Calhoun, J.G. and Maxim, B.R. Predicting Medical School Performance Using the Old and New MCAT. Proceedings of the 20th Annual Conference on Research in Medical Education, Washington, D.C., 135-141, 1981.

Jones, R.F. and Thomae-Forgues, M. A Factor Comparison of Old and New MCAT Scales. Journal of Medical Education, 56:161-166, 1981.

Lord, F.M. and Novick, M.R. Statistical Theories of Mental Test Scores. New York: Addison-Wesley, 1968.

Kerlinger, F.N. and Pedhazur, E.J. Multiple Regression in Behavioral Research. N.Y.: Holt, Reinhart & Winston, 1973.

McGuire, F.L. The New MCAT and Medical Student Performance. Journal of Medical Education, 55:405-408, 1980.

Molider, J.B. and Elstein, A.S. A Factor Analytic Study of the Old and New MCAT Examinations. Proceedings of the 18th Annual Conference on Research in Medical Education, Washington, D.C., 139-144, 1979.

Moiser, C.I. Problems and Designs of Cross-Validation. Educational & Psychological Measurement, 11:5-11, 1951.

Neter, J. and Wasserman, W. Applied Linear Statistical Models. Homewood, Il: R.D. Irwin, 1974.

New Medical College Admission Test Interpretative Manual. Washington, D.C.: Association of American Medical Colleges, 1977.

Sechrest, L. Incremental Validity: A Recommendation. Educational and Psychological Measurement, 23:153-158, 1963.

Wilson, S.R. A Re-Examination of the Relevance of MCAT Science Topics After Five Years. In R.L. Beran (Moderator), Using the MCAT: A Review of Two Issues. GME/GSA Special MCAT Session presented at the meeting of the Association of American Medical Colleges, Washington, D.C., November, 1982.

Wainer, H. and Thissen, D. Graphical Data Analysis. In M.R. Rosenzweig & L. W. Porter (eds.). Annual Review of Psychology, 32:191-241, 1981.